Uniform Error Bounds Involving Logspline Models

By

Charles J. Stone

Technical Report No. 130
November 1987
(revised September 1988)

Research supported in part by NSF Grant DMS-8600409.

Department of Statistics
University of California
Berkeley, California
UNIFORM ERROR BOUNDS INVOLVING LOGSPLINE MODELS

Charles J. Stone

Department of Statistics
University of California, Berkeley

1. INTRODUCTION.

Splines are of increasing importance in statistical theory and methodology. In particular, Stone and Koo (1986) and Stone (1988) considered exponential families of densities in which the logarithm of the density is a spline. Such exponential families are the subject of the present paper, as are corresponding exponential response models. In each context we use an extension of a key result of de Boor (1976) to obtain a bound on the $L_\infty$ norm of the approximation error associated with maximizing the associated expected log-likelihood.

Let $Y$ be a real-valued random variable ranging over a compact interval $I$; without loss of generality, let $I = [0, 1]$. Suppose that $Y$ has a density $f$ that is continuous and positive on $I$.

Let $S$ be a standard vector space of spline functions of a given order $q \geq 1$ on $I$ (piecewise polynomials of degree $q - 1$ or less that are right-continuous on $I$ and continuous at 1) having finite dimension $K \geq 2$. Let $B_1, \ldots, B_K$ be a B-spline basis of $S$ (see de Boor, 1978). Then $B_1, \ldots, B_K$ are nonnegative and sum to 1 on $I$.

Let $\theta_1, \ldots, \theta_K$ be real constants. Set

$$c(\theta_1, \ldots, \theta_K) = \log \left( \int \exp \left( \sum_k \theta_k B_k(y) \right) dy \right)$$

---

1 This research was supported in part by National Science Foundation Grant DMS-8600409.
and

\[ f(y; \theta_1, \ldots, \theta_K) = \exp \left( \sum_k \theta_k B_k(y) - c(\theta_1, \ldots, \theta_K) \right), \quad y \in \mathcal{I}. \]

This defines an exponential family of densities on \( \mathcal{I} \). Observe that, for \( a \in \mathbb{R} \),

\[ c(\theta_1 + a, \ldots, \theta_K + a) = c(\theta_1, \ldots, \theta_K) + a \]

and hence

\[ f(y; \theta_1 + a, \ldots, \theta_K + a) = f(y; \theta_1, \ldots, \theta_K), \quad y \in \mathbb{R}. \]

Consequently the exponential family fails to be identifiable. In order to make it identifiable, we require that \( \theta_K = 0 \).

Let \( \Theta \) denote the collection of ordered \((K - 1)\)-tuples \( \theta_1, \ldots, \theta_{K-1} \) of real numbers. For \( \theta = (\theta_1, \ldots, \theta_{K-1}) \in \Theta \), set

\[ s(y; \theta) = \theta_1 B_1(y) + \cdots + \theta_{K-1} B_{K-1}(y), \quad y \in \mathcal{I}, \]

\[ C(\theta) = \log \left( \int \exp(s(y; \theta))dy \right), \]

and

\[ f(y; \theta) = \exp(s(y; \theta) - C(\theta)), \quad y \in \mathcal{I}. \]

This defines an identifiable exponential family; it is referred to as a \textit{logspline model} since \( \log(f(.; \theta)) \in \mathcal{S} \).

Let \( Y_1 \ldots Y_n \) be independent random variables having common density \( f \), which is not necessarily a member of the indicated logspline model. The corresponding log-likelihood function \( l(\theta) \), \( \theta \in \Theta \), is defined by

\[ l(\theta) = \sum_i \log(f(Y_i; \theta)) = \sum_i [s(Y_i; \theta) - C(\theta)], \quad \theta \in \Theta. \]

Suppose that (for given values of \( Y_1, \ldots, Y_n \)) the log-likelihood function has a maximizing value \( \hat{\theta} \in \Theta \). Then this maximizing value is unique and is called the maximum-likelihood estimate of \( \theta \); the corresponding density \( \hat{f} \) defined by \( \hat{f}(y) = f(y; \hat{\theta}) \) for \( y \in \mathcal{I} \), is referred to as the \textit{logspline density estimate} corresponding to the given logspline model.

The expected log-likelihood function \( \lambda(\theta) \), \( \theta \in \Theta \), is defined by

\[ \lambda(\theta) = \mathbb{E}l(\theta) = n \left[ \int s(y; \theta) f(y)dy - C(\theta) \right], \quad \theta \in \Theta. \]

It follows by a convexity argument that the expected log-likelihood function has a unique maximizing value \( \theta^* \in \Theta \). (Recall that \( f \) is a positive density
on \( \mathcal{I} \) and that \( s(\cdot; \theta) \) is a nonconstant function for \( \theta \neq 0 \). Consider the corresponding density \( Q_s f \) on \( \mathcal{I} \) defined by \( Q_s f(y) = f(y; \theta^*) \), \( y \in \mathcal{I} \). The density \( f \) belongs to the logspline model if and only if \( f = Q_s f \) on \( \mathcal{I} \). When \( f \) does not belong to this model, the function \( f - Q_s f \) plays an important role in the analysis of the asymptotic behavior of the logspline density estimate (see Stone, 1988); roughly speaking, it acts as a bias term.

Given a real-valued function \( g \) on \( \mathcal{I} \), set \( \| g \|_\infty = \sup_{y \in \mathcal{I}} |g(y)| \). Let \( \mathcal{F} \) denote a family of positive densities on \( \mathcal{I} \) such that the family \( \{ \log(f) : f \in \mathcal{F} \} \) is an equicontinuous family. Set

\[
\delta_S(f) = \inf_{s \in S} \| \log(f) - s \|_\infty, \quad f \in \mathcal{F}.
\]

(For an upper bound to \( \delta_S(f) \) in terms of the smoothness of \( \log(f) \), see Theorem XII.1 of de Boor, 1978.) In Section 4 we will obtain an inequality of the form

\[
\| \log(f) - \log(Q_s f) \|_\infty \leq M\delta_S(f), \quad f \in \mathcal{F},
\]

where the positive constant \( M \) depends only on \( \mathcal{F} \), the order of \( S \), and a bound on a suitable “global mesh ratio” of \( S \). The main point of this result is that \( M \) does not depend on \( K = \dim(S) \). It follows from (1) that

\[
\| f - Q_s f \|_\infty \leq \exp(M\delta_S(f) - 1) \| f \|_\infty, \quad f \in \mathcal{F}.
\]

Suppose now that the distribution of \( Y \) depends on a real variable \( x \) that ranges over a compact interval \( \mathcal{I} \); without loss of generality, let \( \mathcal{I} = [0,1] \). Let \( f(\cdot \mid x) \) denote the dependence of density of \( Y \) on \( x \). It is supposed that \( f(y \mid x) \), \( x, y \in \mathcal{I} \), is a continuous and positive function.

Let \( \mathcal{H} \) be a standard finite-dimensional vector space of spline functions of a given order on \( \mathcal{I} \) having dimension \( J \geq 1 \), and let \( H_1, \ldots, H_J \) be a \( B \)-spline basis of \( \mathcal{H} \).

Let \( B \) denote the collection of \( J \times (K - 1) \) matrices \( \beta = (\beta_{jk}) \) of real numbers \( \beta_{jk} \), \( 1 \leq j \leq J \) and \( 1 \leq k \leq K - 1 \). Let \( \beta \in B \). For \( 1 \leq k \leq K - 1 \), let \( h_k(\cdot; \beta) \) be the real-valued function on \( \mathcal{I} \) defined by

\[
h_k(x; \beta) = \sum_j \beta_{jk} H_j(x), \quad x \in \mathcal{I}.
\]

Set

\[
h(x; \beta) = (h_1(x; \beta), \ldots, h_{K-1}(x; \beta)), \quad x \in \mathcal{I}.
\]

Then \( h(\cdot; \beta) \) is an \( \mathbb{R}^{K-1} \)-valued function on \( \mathcal{I} \).

The logspline response model corresponding to \( \mathcal{H} \) and \( S \) is defined by

\[
f(y \mid x; \beta) = f(y; h(x; \beta)) = \exp(s(y; h(x; \beta)) - C(h(x; \beta)))
\]
for $\beta \in B$ and $x, y \in \mathcal{I}$. Observe that, for $\beta \in B$ and $x \in \mathcal{I}$, $f(\cdot | x; \beta)$ is a positive density on $\mathcal{I}$.

Let $x_1, \ldots, x_n \in \mathcal{I}$ and let $Y_1, \ldots, Y_n$ be independent random variables such that $Y_i$ has density $f(\cdot | x_i)$. The corresponding log-likelihood function $l(\beta)$, $\beta \in B$, is defined by

$$l(\beta) = \sum_i \log(f(Y_i | x_i; \beta)) = \sum_i (s(Y_i; h(x_i; \beta)) - C(h(x_i; \beta))), \quad \beta \in B.$$ 

The expected log-likelihood function $\lambda(\beta)$, $\beta \in B$, is defined by

$$\lambda(\beta) = E[l(\beta)] = \sum_i \left[ \int s(y; h(x_i; \beta))f(y | x_i)dy - C(h(x_i; \beta)) \right], \quad \beta \in B.$$ 

Suppose that $\mathcal{H}$ is identifiable from $x_1, \ldots, x_n$; that is, that if $h \in \mathcal{H}$ and $h(x_1) = \cdots = h(x_n) = 0$, then $h = 0$ on $\mathcal{I}$. Then, by a convexity argument, the expected log-likelihood function has a unique maximum $\beta^* \in B$. Consider the corresponding function $Q_{s f}$ on $\mathcal{I} \times \mathcal{I}$ defined by

$$Q_{s f}(y | x) = f(y | x; \beta^*), \quad x, y \in \mathcal{I}.$$ 

Let $T$ denote the tensor product of $\mathcal{H}$ and $\mathcal{S}$; that is, the vector space of real-valued functions on $\mathcal{I} \times \mathcal{I}$ spanned by functions of the form $h(x)s(y)$, $x, y \in \mathcal{I}$, as $h$ and $s$ range over $\mathcal{H}$ and $\mathcal{S}$ respectively. Then $T$ has dimension $JK$, and the functions $H_j(x)B_k(y)$, $x, y \in \mathcal{I}$, $1 \leq j \leq J$ and $1 \leq k \leq K$ form a basis of $T$.

Given a real-valued function $g$ on $\mathcal{I} \times \mathcal{I}$, set $\| g \|_\infty = \sup_{\mathcal{I} \times \mathcal{I}} g(x, y)$. Let $\mathcal{F}$ denote a family of continuous and positive functions $f$ on $\mathcal{I} \times \mathcal{I}$ such that $f(\cdot | x)$ is a density on $\mathcal{I}$ for $x \in \mathcal{I}$ and $\{\log(f) : f \in \mathcal{F}\}$ is an equicontinuous family of functions on $\mathcal{I} \times \mathcal{I}$. Set

$$\delta_T(f) = \inf_{t \in \mathcal{F}} \| \log(f) - t \|_\infty, \quad f \in \mathcal{F}.$$ 

(For an upper bound to $\delta_T(f)$ in terms of the smoothness of $\log(f)$, see Theorem 12.8 of Schumaker, 1981.) In Section 5 we will obtain an inequality of the form

$$(2) \quad \| \log(f) - \log(Q_T f) \|_\infty \leq M \delta_T(f), \quad f \in \mathcal{F},$$

where the positive constant $M$ depends on $\mathcal{F}$, the orders of $\mathcal{H}$ and $\mathcal{S}$, bounds on the global mesh ratios of $\mathcal{H}$ and $\mathcal{S}$, and a measure of regularity of $x_1, \ldots, x_n$ that depends on $\mathcal{H}$. The main point of this result is that $M$ does not depend on $J = \dim(\mathcal{H})$ or $K = \dim(\mathcal{S})$. 

4
2. PRELIMINARY INEQUALITIES

The bound on the global mesh ratio for $S$ described in de Boor (1976) is equivalent to a bound of the form

\[ M^{-1}K^{-1} \leq \int B_k(y)dy \leq M_1K^{-1}, \quad 1 \leq k \leq K, \]

where $M_1 > 1$ is a constant. Since the support of $B_k$ is an interval having length $q \int B_k(y)dy$, where $q$ is the order of $S$, (3) can be written as a two-sided bound on this length. Under (3) there is a constant $M_2 > 1$ (depending on the order of $S$) such that, for $\theta_1, \ldots, \theta_K \in \mathbb{R}$,

\[ M_1^{-1}M_2^{-1}K^{-1} \sum_k \theta_k^2 \leq \int \left( \sum_k \theta_k B_k(y) \right)^2 dy \leq M_1K^{-1} \sum_k \theta_k^2 \]

(see (7) of de Boor, 1976).

Similarly, we assume that

\[ M_1^{-1}J^{-1} \leq \int H_j(x)dx \leq M_1J^{-1}, \quad 1 \leq j \leq J. \]

Under (5) it can be assumed that, for $\beta_1, \ldots, \beta_J \in \mathbb{R}$,

\[ M_1^{-1}M_2^{-1}J^{-1} \sum_j \beta_j^2 \leq \int \left( \sum_j \beta_j H_j(x) \right)^2 dx \leq M_1J^{-1} \sum_j \beta_j^2. \]

For a given order $q$ of $H$, the functions in $H$ are piecewise polynomials of degree $q - 1$ or less. In light of (5), a natural regularity assumption on $x_1, \ldots, x_n$ is that

\[ M_3^{-1}n \int h^2(x)dx \leq \sum_i h^2(x_i) \leq M_3n \int h^2(x)dx, \quad h \in H, \]

for some constant $M_3 > 1$. It follows from (7) that $H$ is identifiable from $x_1, \ldots, x_n$. It also follows from (7), by choosing $M_3$ larger if necessary depending on the order of $H$, that

\[ \sum_i H_j(x_i) \leq M_3J^{-1}n, \quad 1 \leq j \leq J. \]

(Let $h$ denote the sum of the $H_k$'s whose support overlaps with that of $H_j$; note that $H_j \leq 1 = h = h^2$ on the support of $H_j$.)
Let $\rho$ be a positive (Borel) function on $I$ such that, for some constant $M_4 > 1$,

$$M_4^{-1} \leq \rho(y) \leq M_4, \quad y \in I.$$  

(9)

For the real-valued function $g$ on $I \times I$, let $\| g \|_2$ be the nonnegative square root of

$$\| g \|^2 = \sum_i \int g^2(x_i, y)\rho(y)dy.$$  

For $1 \leq j \leq J$ and $1 \leq k \leq K$, define $B_{jk}$ on $I \times I$ by

$$B_{jk}(x, y) = H_j(x)B_k(y), \quad x, y \in I.$$  

It follows from (4), (6), (7) and (9) that, for $\beta \in B$,

$$\frac{1}{M_1^2 M_2 M_3 M_4 JK} \sum_j \sum_k \beta_{jk}^2 \leq \left\| \sum_j \sum_k \beta_{jk}B_{jk} \right\|^2_2 \leq \frac{M_1^2 M_3 M_4 n}{JK} \sum_j \sum_k \beta_{jk}^2.$$  

(10)

3. THE INVERSE GRAM MATRIX

Consider the $K \times K$ matrix $M$ whose $(k, l)$th entry is $\int B_k(y)B_l(y)\rho(y)dy$. It follows from (4) that $M$ is invertible. Let $\alpha_{kl}$ denote the $(k, l)$th entry of $M^{-1}$. Then

$$\| M^{-1} \|_\infty \leq \max_l \sum_k |\alpha_{kl}|.$$  

By a slight extension of a result in de Boor (1976), there is a constant $M_8 > 1$, depending on $M_1$, $M_2$ and $M_4$, such that

$$\| M^{-1} \|_\infty \leq M_8 K$$  

(11)

(see the proof of (18) below). This has the following consequence.

**Lemma 1.** Set $g = \sum k \theta_k B_k$. Then

$$\max_k |\theta_k| \leq M_8 K \max_k \left| \int g(y)B_k(y)\rho(y)dy \right|.$$  

For real-valued functions $g_1$ and $g_2$ on $I \times I$ such that the norms $\| g_1 \|_2$ and $\| g_2 \|_2$ are finite, set

$$
\langle g_1, g_2 \rangle = \sum_i \int g_1(x_i, y)g_2(x_i, y)\rho(y)dy.
$$

Then $\| g \|_2^2 = \langle g, g \rangle$. Consider now the $JK \times JK$ matrix $M$ whose $((j,k),(l,m))$th entry is the inner product $\langle B_{jk}, B_{lm} \rangle$ of $B_{jk}$ and $B_{lm}$. It follows from (10) that $M$ is invertible. Let $\alpha_{jklm}$ denote the $((j,k),(l,m))$th entry of $M^{-1}$. Then

$$
\| M^{-1} \|_\infty = \max_{j,k} \sum_l \sum_m |\alpha_{jklm}|.
$$

We will now imitate the elegant proof of (11) above in de Boor's paper (see also Descloux, 1972).

Set $f_{jk} = \sum_l \sum_m \alpha_{jklm}B_{lm}$. Then $\langle f_{jk}, B_{lm} \rangle$ equals 1 if $j = l$ and $k = m$ and it equals zero otherwise. Consequently,

$$
0 < \| f_{jk} \|_2^2 = \alpha_{jklk}.
$$

Set $M_5 = M_1^2M_2^2M_3M_4 > 1$. Then, by (10),

$$
M_5^{-1}J^{-1}K^{-1}n\alpha_{jklk}^2 \leq M_5^{-1}J^{-1}K^{-1}n\sum_l \sum_m \alpha_{jklm}^2 \leq \| f_{jk} \|_2^2 = \alpha_{jklk}.
$$

Therefore

$$
\alpha_{jklk} \leq M_5JKn^{-1}
$$

and

$$
\sum_l \sum_m \alpha_{jklm}^2 \leq M_5JKn^{-1}\alpha_{jklk} \leq (M_5JKn^{-1})^2.
$$

Set $M_6 = M_1^2M_2M_3M_4 > 1$.

**Lemma 2.** There is a constant $M_7 > 1$, depending on $M_6$, such that

$$
|\alpha_{jklm}| \leq M_5M_6M_7JKM_7^{-(|j - l| + |k - m|)}n^{-1}.
$$

**Proof.** Let $(j,k)$ be given and let $v, w \in \mathbb{R}$ with $v^2 + w^2 = 1$. For $c \in \mathbb{R}$, set

$$
S_c = \{(l,m) : v(l - j) + w(m - k) \geq c\}$$

$\mathbb{R}$.
and
\[ g_c = \sum_{S_c} \sum_{S_c} \alpha_{jklm} B_{lm}. \]

Let \( c > 0 \). Since \( f_{jk} \) is orthogonal to \( B_{lm} \) for \( (l, m) \neq (j, k) \), \( g_c \) is orthogonal to \( f_{jk} \). There is a positive constant \( u \), depending only on the order of \( \mathcal{H} \) and \( \mathcal{S} \), such that if \( (l, m) \in S_c \) and \( (l_1, m_1) \neq S_{c-u} \), then \( B_{lm} \) and \( B_{l_1m_1} \) have disjoint support and hence are orthogonal to each other. Consequently, \( g_c \) is orthogonal to \( f_{jk} - g_{c-u} \) and hence to \( g_{c-u} \). Therefore,
\[ ||g_{c-u}||_2^2 + ||g_c||_2^2 = ||g_{c-u} - g_c||_2^2 \]
and hence
\[ ||g_{c-u}||_2^2 \leq ||g_{c-u} - g_u||_2^2. \]

Now
\[ g_{c-u} - g_u = \sum_{S_{c-u,c}} \sum_{S_{c-u,c}} \alpha_{jklm} B_{lm}, \]
where
\[ S_{c-u,c} = S_{c-u} \setminus S_c = \{(l, m) : c - u \leq v(l - j) + w(m - k) < c\}. \]

We conclude from (10) and (14) that
\[ \sum_{S_{c-u,c}} \sum_{S_{c-u,c}} \alpha_{jklm}^2 \geq M_6^{-2} \sum_{S_{c-u}} \sum_{S_{c-u}} \alpha_{jklm}^2, \quad c > 0. \]

Set
\[ a_\nu = \sum_{S_{c+(\nu-1)u,c+\nu u}} \alpha_{jklm}^2, \quad \nu = 0, 1, 2, \ldots. \]

By (15),
\[ |a_\nu| \geq M_6^{-2}(|a_\nu| + |a_{\nu+1}| + \cdots), \quad \nu = 0, 1, 2, \ldots. \]

According to Lemma 2 of de Boor (1976), (16) implies that
\[ |a_\nu| \leq |a_0| M_6^2 (1 - M_6^{-2})^\nu, \quad \nu = 0, 1, 2, \ldots. \]

By (13) and (17),
\[ |a_\nu| \leq (M_5 M_6 J K n^{-1})^2 (1 - M_6^{-2})^\nu, \quad \nu = 0, 1, 2, \ldots. \]

It follows by choosing \( v, w, \) and \( c \) appropriately that if
\[ \nu \leq u^{-1}[(l - j)^2 + (m - k)^2]^{1/2}, \]
then
\[ |\alpha_{jklm}| \leq M_5 M_6 J K (1 - M_6^{-2})^{\nu/2} n^{-1}. \]
This yields the conclusion of the lemma.

Set

\[ M_8 = M_5 M_6 M_7 (M_7 + 1)^2 (M_7 - 1)^{-2} > 1. \]

It follows from (12) and Lemma 2 that

(18) \[ \| M^{-1} \|_\infty \leq M_8 J Kn^{-1}. \]

This inequality has the following implication.

**Lemma 3.** Set

\[ g = \sum_j \sum_k \beta_{jk} B_{jk}. \]

Then

\[ \max_{j,k} |\beta_{jk}| \leq M_8 J Kn^{-1} \max_{j,k} |\langle g, B_{jk} \rangle|. \]

**4. LOGSPIINE MODELS**

In this section, we obtain (1). For \( f \) a positive density on \( I \) and \( 0 < a < 1 \), let \( f_a \) denote the density on \( I \) defined by

\[ f_a(y) = \frac{f^a(y)}{\int f^a(y) dy}. \]

It can be assumed that \( f_a \in \mathcal{F} \) for \( f \in \mathcal{F} \) and \( 0 < a < 1 \). (Extend \( \mathcal{F} \) if necessary.)

Choose \( s \in \mathcal{S} \) and define the real-valued function \( g \) on \( \mathbb{R} \) by

\[ \int \exp(t s(y) - g(t)) Q_s f(y) dy = 1. \]

Then

\[ g'(0) = \int s(y) Q_s f(y) dy. \]

Also

\[ \int [\log(Q_s f(y)) + t s(y) - g(t)] f(y) dy \]

is maximized at \( t = 0 \); hence

\[ g'(0) = \int s(y) f(y) dy. \]
Thus
\[
\int s(y)[Qsf(y) - f(y)]dy = 0.
\]
Consequently,
\[
\int B_k(y)[Qsf(y) - f(y)]dy = 0, \quad 1 \leq k \leq K,
\]
or, equivalently,
\[
\int B_k(y)[Qsf(y) - f(y)]dy = 0, \quad 1 \leq k \leq K - 1.
\]
Formula (20) can also be written as
\[
\frac{\partial C}{\partial \theta_k}(\theta^*) = \int B_k(y)f(y)dy, \quad 1 \leq k \leq K - 1.
\]
Let \( K \) be a fixed positive integer and let \( S \) otherwise vary subject to (3). Then \( B_1 \ldots B_K \) depend continuously (in the \( L_2 \) norm) on the knot sequence defining \( S \). Thus it follows from (21) and the properties of the Hessian matrix of \( C(\cdot) \) (e.g., it is negative definite) that \( \theta^* \) depends continuously on \( \int B_k(y)f(y)dy, 1 \leq k \leq K - 1, \) and the knot sequence defining \( f \).
Let \( f \in F \). There is an \( s \in S \) such that \( \| \log(f) - s \|_\infty = \delta_S(f) \). Since \( f \) is a density on \( I \), we conclude that
\[
\left| \log \left( \int \exp(s(y))dy \right) \right| \leq \delta_S(f).
\]
Consequently, there is a \( \tilde{\theta} \in \Theta \) such that
\[
\| \log(f) - \log(f(\cdot; \tilde{\theta})) \|_\infty \leq 2\delta_S(f).
\]
Note that \( Qsf \tilde{f} = \tilde{f} \), where \( \tilde{f} = f(\cdot; \tilde{\theta}) \). Thus it follows from (22) and the continuity properties of \( \theta^* \) described above that there is a positive constant \( M_{1K} \) (depending on \( M_1 \) and \( F \) as well as \( K \)) such that
\[
\| \log(f(\cdot; \theta^*)) - \log(f(\cdot; \tilde{\theta})) \|_\infty \leq M_{1K}\delta_S(f)
\]
and hence
\[
\| \log(f) - \log(Qsf) \|_\infty \leq (M_{1K} + 2)\delta_S(f), \quad f \in F.
\]
Choose \( \bar{\theta} \in \Theta \) such that (22) holds and set \( \bar{f} = f(\cdot; \bar{\theta}) \). Then
\[
\| \log(f) - \log(\bar{f}) \|_\infty \leq 2\delta_S(f).
\]
There are constants $M_9, M_{10} > 1$, depending on $F$, such that

\[(25) \quad \| f - \bar{f} \|_\infty \leq M_9 \delta_s(F)\]

and

\[(26) \quad M_{10}^{-1} \leq \bar{f}(y) \leq M_{10}, \quad y \in \mathcal{T}.\]

By (3), (19) and (25),

\[(27) \quad \int B_k(y)[Q_f(y) - \bar{f}(y)]dy \leq M_1M_9K^{-1} \delta_s(f), \quad 1 \leq k \leq K.\]

Write

\[\log(Qsf) - \log(\bar{f}) = \sum_k \theta_k B_k\]

and set $\epsilon = \max_k |\theta_k|$. Now $\| \log(Qsf) - \log(\bar{f}) \|_\infty \leq \epsilon$ and hence

\[(28) \quad \| \log(f) - \log(Qsf) \|_\infty \leq \epsilon + 2 \delta_s(f).\]

It follows from (viii) on Page 155 of de Boor (1978) that there is a positive constant $M_{11}$, depending on the order of $S$, such that

\[(29) \quad \epsilon \leq M_{11} \| \log(Qsf) - \log(\bar{f}) \|_\infty.\]

Suppose that $\epsilon \leq 1$. Since $Qsf = \bar{f} \exp(\sum_k \theta_k B_k)$, we conclude from (26) that

\[\| Qsf - \bar{f} - \bar{f} \sum_k \theta_k B_k \|_\infty \leq M_{10}\epsilon^2\]

and hence from (3) and (27) that, for $1 \leq k \leq K$,

\[(30) \quad \left| \int B_k(y) \sum_i \theta_i B_i(y) \bar{f}(y)dy \right| \leq M_1M_9K^{-1} \delta_s(f) + M_1M_{10}K^{-1}\epsilon^2.\]

According to (26), (30) and Lemma 1, there is a constant $M_{12} > 1$, depending on $M_1$, $M_2$ and $M_{10}$, such that

\[\epsilon \leq M_1M_9M_{12} \delta_s(f) + M_1M_{10}M_{12}\epsilon^2.\]

Suppose now that

\[(31) \quad M_1M_{10}M_{12}\epsilon \leq \frac{1}{2}.\]

Then $\epsilon \leq 2M_1M_9M_{12}\delta_s(f)$ and hence, by (28),

\[(32) \quad \| \log(f) - \log(Qsf) \|_\infty \leq M_{13}\delta_s(f),\]

11
where $M_{13} = 2(M_1M_9M_{12} + 1)$. According to (29), a sufficient condition for (31) and hence for (32) is

$$\| \log(Qs f) - \log(\bar{f}) \|_\infty \leq M_{14}^{-1},$$

where $M_{14} = 2M_1M_9M_{11}M_{12}$.

Let $0 < \delta < 2^{-1}M_{13}^{-1}M_{14}^{-1}$.

There is a positive integer $K_0$, depending on $M_1$ and the order of $S$, such that

$$\delta_S(f) \leq \delta, \quad K \geq K_0 \text{ and } f \in \mathcal{F}$$

(see Page 167 of de Boor, 1978). Let $K \geq K_0$. Suppose that

$$\| \log(f) - \log(Qs f) \|_\infty \leq 2^{-1}M_{14}^{-1}.$$ 

Then (33) follows from (24), so (32) holds.

We will now verify that (35) necessarily holds for $K \geq K_0$. Suppose not. Now

$$\| \log(f_a) - \log(Qs f_a) \|_\infty$$

is continuous in $a$ for $0 < a < 1$ and it approaches 0 as $a \to 0$. (According to an earlier argument, $\theta^*$ is continuous in $a$.) Thus there is a value of $a \in (0,1)$ such that

$$\| \log(f_a) - \log(Qs f_a) \|_\infty = 2^{-1}M_{14}^{-1}.$$ 

By the previous argument, (32) and (34) hold with $f$ replaced by $f_a$; hence

$$\| \log(f_a) - \log(Qs f_a) \|_\infty \leq M_{13}\delta_S(f_a) \leq M_{13}\delta < 2^{-1}M_{14}^{-1},$$

which yields a contradiction.

We have now shown that

$$\| \log(f) - \log(Qs f) \|_\infty \leq M_{13}\delta_S(f), \quad K \geq K_0 \text{ and } f \in \mathcal{F}.$$ 

The desired inequality (1) follows from (36) together with (23) for $1 \leq K < K_0$. 

12
5. LOGSPLINE RESPONSE MODELS

In this section, we obtain (2). For \( f \) a positive function on \( I \times I \) such that \( f(\cdot \mid x) \) is a density on \( I \) for each \( x \in I \) and for \( 0 < a < 1 \), let \( f_a \) be defined on \( I \times I \) by

\[
f_a(y \mid x) = \frac{f^a(y \mid x)}{\int f^a(y \mid x)dy}.
\]

It can be assumed that \( f_a \in \mathcal{F} \) for \( f \in \mathcal{F} \). (Extend \( \mathcal{F} \) if necessary.)

Let \( 1 \leq k \leq K - 1 \). Choose \( h \in \mathcal{H} \) and let \( h \) be the \( \mathbb{R}^{K-1} \)-valued function on \( I \) whose \( k \)th component is \( h \) and whose other components are zero. Define the real-valued function \( g \) on \( \mathbb{R} \) by

\[
g(t) = \sum_i \left[ \int s(y; h(x_i; \beta^*)) + t h(x_i)f(y \mid x_i)dy - C(h(x_i; \beta^*) + t h(x_i)) \right].
\]

Then

\[
0 = g'(0) = \sum_i h(x_i) \left[ \int B_k(y)f(y \mid x_i)dy - \frac{\partial C}{\partial \theta_k}(h(x_i; \beta^*)) \right].
\]

Thus, for \( 1 \leq j \leq J \) and \( 1 \leq k \leq K - 1 \),

\[
(37) \quad \sum_i H_j(x_i) \frac{\partial C}{\partial \theta_k}(h(x_i; \beta^*)) = \sum_i H_j(x_i) \int B_k(y)f(y \mid x_i)dy,
\]

which can also be written as

\[
\sum_i H_j(x_i) \int B_k(y)[f(y \mid x_i) - Q_T f(y \mid x_i)]dy = 0
\]

or, equivalently, as

\[
(38) \quad \sum_i H_j(x_i) \int B_k(y)[f(y \mid x_i) - Q_T f(y \mid x_i)]dy = 0.
\]

Let \( f \in \mathcal{F} \). There is a \( t \in \mathcal{T} \) such that \( \| \log(f) - t \|_\infty = \delta_T(f) \). Let \( x \in I \). Since \( f(\cdot \mid x) \) is a density on \( I \), we conclude that

\[
\left| \log \left( \int e^{t(x,y)}dy \right) \right| \leq \delta_T(f), \quad x \in I.
\]

Consequently, there is a \( \tilde{\beta} \in \mathcal{B} \) such that

\[
(39) \quad \| \log(f) - \log(f(\cdot \mid \cdot) ; \tilde{\beta}) \|_\infty \leq 2\delta_T(f).
\]
Let $J$ and $K$ be fixed positive integers and let $\mathcal{H}, \mathcal{S}$ and $x_1 \ldots x_n$ otherwise vary subject to (3), (5) and (7). It follows from (37) that there is a positive constant $M_{JK}$ (depending on $M_1, M_3$ and $\mathcal{F}$ as well as $J$ and $K$) such that

$$\| \log(f(\cdot | \cdot; \beta^*) - \log(f(\cdot | \cdot; \hat{\beta})) \|_{\infty} \leq M_{JK} \delta_T(f).$$

We conclude from (39) and (40) that

$$\| \log(f) - \log(Q_T f) \|_{\infty} \leq (M_{JK} + 2) \delta_T(f), \quad f \in \mathcal{F}.$$

There are positive integers $J_0$ and $K_0$ and there is a positive constant $M_9$, depending on $\mathcal{F}, M_1 \ldots M_4$ and the orders of $\mathcal{H}$ and $\mathcal{S}$ such that

$$\| \log(f) - \log(Q_T f) \|_{\infty} \leq M_9 \delta_T(f), \quad J \geq J_0, \quad K \geq K_0 \quad \text{and} \quad f \in \mathcal{F}.$$  

The argument used to prove (42) is a refinement of that used to prove (36). To start off, choose $\bar{t} \in T$ such that $\| \log(f) - \bar{t} \|_{\infty} = \delta_T(f)$, set

$$\bar{c}(x) = \log \left( \int \exp(\bar{t}(x, y)) dy \right), \quad x \in \mathcal{I},$$

and note that

$$| \bar{c}(x) | \leq \delta_T(f), \quad x \in \mathcal{I}.$$  

Define $\bar{f}$ on $\mathcal{I} \times \mathcal{I}$ by $\bar{f}(y | x) = \exp(\bar{t}(x, y) - \bar{c}(x))$. Then

$$\| \log(f) - \log(\bar{f}) \|_{\infty} \leq 2 \delta_T(f).$$  

There are constants $M_{10}, M_{11} > 1$, depending on $\mathcal{F}$, such that

$$\| f - \bar{f} \|_{\infty} \leq M_{10} \delta_T(f)$$

and

$$M_{11}^{-1} \leq \bar{f}(y | x) \leq M_{11}, \quad x, y \in \mathcal{I}.$$  

By (3), (8), (38) and (43),

$$\left| \sum_i H_j(x_i) \int B_k(y)[Q_T f(y | x_i) - \bar{f}(y | x_i)] dy \right| \leq \frac{M_1 M_3 M_{10}}{JK} n \delta_T(f)$$

for $1 \leq j \leq J$ and $1 \leq k \leq K$.

Write

$$\log(Q_T f(y | x)) = t^*(x, y) - c^*(x), \quad x, y \in \mathcal{I},$$

where $t^* \in T$, and set $t = t^* - \bar{t}$. Then

$$Q_T f(y | x) = \exp(t(x, y) + \bar{c}(x) - c^*(x)) \bar{f}(y | x), \quad x, y \in \mathcal{I},$$
\[ c^*(x) = \log \left( \int \exp(t(x, y) + \bar{c}(x)) \bar{f}(y \mid x) dy \right) = \log \left( (1 + \int \exp(t(x, y) + \bar{c}(x)) - 1) f(y \mid x) dy \right) \]

for \( x \in \mathcal{I} \), and

\[ Q_T f(y \mid x) - \bar{f}(y \mid x) = [\exp(t(x, y) + \bar{c}(x) - c^*(x)) - 1] f(y \mid x), \quad x, y \in \mathcal{I}. \]

Thus

\[ c^*(x) - \bar{c}(x) \approx \int t(x, y) \bar{f}(y \mid x) dy, \quad x \in \mathcal{I}, \]

and hence

\[ Q_T f(y \mid x) - \bar{f}(y \mid x) \approx \left[ t(x, y) - \int t(x, y) \bar{f}(y \mid x) dy \right] \bar{f}(y \mid x) \]

for \( x, y \in \mathcal{I} \).

Write

\[ t(x, y) = \sum_j \sum_k \beta_{jk} H_j(x) B_k(y), \quad x, y \in \mathcal{I}. \]

It follows by a double application of (viii) on Page 155 of de Boor (1978) that there is a positive constant \( M_{12} \), depending on the order of \( \mathcal{H} \) and \( \mathcal{S} \), such that

\[ \max_{j, k} | \beta_{jk} | \leq M_{12} \| t \|_\infty. \]

Choose \( \eta > 0 \). Now

\[ \int t(x, y) \bar{f}(y \mid x) dy = \sum_k \int B_k(y) \sum_j \beta_{jk} H_j(x) \bar{f}(y \mid x) dy. \]

Choose \( x_j \) in the support of \( H_j \). Define \( h \in \mathcal{H} \) by

\[ h(x) = \sum_k \int B_k(y) \sum_j \beta_{jk} H_j(x) \bar{f}(y \mid x_j) dy = \sum_j H_j(x) \sum_k \beta_{jk} \int B_k(y) \bar{f}(y \mid x_j) dy. \]

There is a positive integer \( J_0 \), depending on \( M_1, M_{12} \) and \( \mathcal{F} \) such that

\[ \left| \int t(x, y) \bar{f}(y \mid x) dy - h(x) \right| \leq \eta \| t \|_\infty, \quad J \geq J_0 \text{ and } x \in \mathcal{I}. \]
After replacing \( t^*(x, y) \) by \( t^*(x, y) - h(x) \) and replacing \( c^*(x) \) by \( c^*(x) - h(x) \), we have that

\[
(45) \quad \left| \int t(x, y) f(y \mid x) \, dy \right| \leq \eta \| t \|_{\infty}, \quad J \geq J_0 \text{ and } x \in \mathcal{I}.
\]

The argument used to prove (42) from (44) and (45) is similar to that used to prove (36), except that Lemma 3 is used instead of Lemma 1 and Theorem 12.8 of Schumaker (1981) is used instead of Page 167 of de Boor (1978).

Next it will be shown that, for each positive integer \( K \), there is a positive integer \( J_0 \) and there is a positive constant \( M_{13} \), both depending on \( \mathcal{F} \), \( M_1, \ldots, M_4 \) and the order of \( \mathcal{H} \) and \( \mathcal{S} \), such that

\[
(46) \quad \| \log(f) - \log(Q_T f) \|_{\infty} \leq M_{13} \delta_\varphi(f), \quad J \geq J_0 \text{ and } f \in \mathcal{F}.
\]

To this end, write

\[
Q_T f(x \mid y) = \exp \left( \sum_k \theta_k(x) B_k(y) - c(x) \right), \quad x, y \in \mathcal{I}.
\]

From (21) we conclude that (as \( f \) varies over \( \mathcal{F} \), etc.) the resulting functions \( \theta_k(\cdot), 1 \leq k \leq K - 1 \), are uniformly bounded and equicontinuous, and there is a positive constant \( M_{14} \) such that

\[
(47) \quad \max_{1 \leq k \leq K - 1} \delta_\mathcal{H}(\theta_k(\cdot)) \leq M_{14} \delta_\mathcal{T}(f).
\]

Observe that

\[
\max_{1 \leq k \leq K - 1} \delta_\mathcal{H}(\theta_k(\cdot))
\]

can be made arbitrary small by making \( J \) sufficiently large (see Page 167 of de Boor, 1978). According to (1), there is a positive constant \( M_{15} \) such that

\[
(48) \quad \left| \log(f(y \mid x)) - \left( \sum_k \theta_k(x) B_k(y) - c(x) \right) \right| \leq M_{15} \delta_\mathcal{T}(f), \quad x, y \in \mathcal{I}.
\]

It follows from (19) that

\[
\int B_k(y) \left[ \exp \left( \sum_m \theta_m(x) B_m(y) - c(x) \right) - f(y \mid x) \right] \, dy = 0
\]

for \( x \in \mathcal{I} \) and \( 1 \leq k \leq K \) and hence that

\[
\sum_i H_j(x_i) \int B_k(y) \left[ \exp \left( \sum_m \theta_m(x_i) B_m(y) - c(x_i) \right) - f(y \mid x_i) \right] \, dy = 0
\]
for $1 \leq j \leq J$ and $1 \leq k \leq K$. Thus we conclude from (38) that

$$\sum_k H_j(x_i) \int B_k(y) \left[ \exp \left( \sum_m \theta_m(x_i) B_m(y) - c(x_i) \right) - Q_T f(y \mid x_i) \right] \, dy = 0$$

for $1 \leq j \leq J$ and $1 \leq k \leq K$.

For $1 \leq k \leq K - 1$, choose $\tilde{h}_k \in \mathcal{H}$ such that

$$| \theta_k(x) - \tilde{h}_k(x) | = \delta_\mathcal{H}(\theta_k(\cdot)), \quad x \in \mathcal{I}.$$

Set

$$\bar{c}(x) = \log \left( \int \exp \left( \sum_k \tilde{h}_k(x) B_k(y) \right) \, dy \right), \quad x \in \mathcal{I},$$

and define $\bar{f}$ on $\mathcal{I} \times \mathcal{I}$ by

$$\bar{f}(y \mid x) = \exp \left( \sum_k \tilde{h}_k(x) B_k(y) - \bar{c}(x) \right).$$

Write

$$Q_T f(y \mid x) = \exp \left( \sum_k h^*(x) B_k(y) - \bar{c}^*(x) \right), \quad x, y \in \mathcal{I},$$

where $h^* \in \mathcal{H}$ for $1 \leq k \leq K - 1$. It now follows by arguing as in the proofs of (36) and (42) that there is a positive constant $M_{16}$ such that

$$| \theta_k(x) - h^*(x) | \leq M_{16} \max_{1 \leq k \leq K-1} \delta_\mathcal{H}(\theta_k(\cdot)), \quad 1 \leq k \leq K - 1 \text{ and } x \in \mathcal{I}.$$

Thus there is a positive constant $M_{17}$ such that

$$\left| \log(Q_T f(y \mid x)) - \left( \sum_k \theta_k(x) B_k(y) - c(x) \right) \right| \leq M_{17} \max_{1 \leq k \leq K-1} \delta_\mathcal{H}(\theta_k(\cdot)). \quad (49)$$

The desired result (46) follows from (47)-(49).

Finally it will be shown that, for each positive integer $J$, there is a positive integer $K_0$ and there is a positive constant $M_{18}$, both depending on $\mathcal{F}$, $M_1, \ldots, M_4$ and the order of $\mathcal{H}$ and $\mathcal{S}$, such that

$$\| \log(f) - \log(Q_T f) \|_\infty \leq M_{18} \beta_T(f), \quad K \geq K_0 \text{ and } f \in \mathcal{F}. \quad (50)$$

To this end, let $\beta_1(\cdot), \ldots, \beta_J(\cdot)$ be the real-valued functions on $\mathcal{I}$ such that

$$\sum_i \left[ \log(f(y \mid x_i)) - \sum_j \beta_j(y) H_j(x_i) \right]^2$$
minimizes
\[ \sum_i \left[ \log(f(y \mid x_i)) - \sum_j \beta_j H_j(x_i) \right]^2 \]
for \( y \in I \). It follows from the appropriate analog of Lemma 2 that, as \( f \) varies over \( F \), etc., the resulting functions \( \beta_1(\cdot), \ldots, \beta_J(\cdot) \) are uniformly bounded and equicontinuous, that there is a positive constant \( M_{19} \) such that
\[ \max_{1 \leq j \leq J} \delta_S(\beta_j(\cdot)) \leq M_{19} \delta_T(f), \]
and that there is a positive constant \( M_{20} \) such that
\[ \left| \log(f(y \mid x)) - \sum_j \beta_j(y) H_j(x) \right| \leq M_{20} \delta_T(f), \quad x, y \in I. \]

Observe that
\[ \max_{1 \leq j \leq J} \delta_S(\beta_j(\cdot)) \]
can be made arbitrarily small by making \( K \) sufficiently large. For \( 1 \leq j \leq J \) choose \( \bar{s}_j \in S \) such that
\[ |\beta_j(y) - \bar{s}_j(y)| = \delta_S(\beta_j(\cdot)), \quad y \in I. \]
Set
\[ \bar{c}(x) = \log \left( \int \exp \left( \sum_j H_j(x) \bar{s}_j(y) dy \right) \right), \quad x \in I. \]

There is a constant \( M_{21} \) such that
\[ |\bar{c}(x)| \leq M_{21} \delta_T(f), \quad x \in I. \]
Define \( f \) on \( I \times I \) by \( \bar{f}(y \mid x) = \exp(\sum_j H_j(x) \bar{s}_j(y) - \bar{c}(x)) \). Write
\[ Q_T f(y \mid x) = \exp \left( \sum_j H_j(x) s^*_j(y) - c'(x) \right), \quad x, y \in I, \]
where \( s^*_j \in S \) for \( 1 \leq j \leq J \). It follows as in the proofs of (36), (42) and (49) that there is a positive constant \( M_{22} \) such that
\[ |\log(Q_T f(y \mid x)) - \log(\bar{f}(y \mid x))| \leq M_{22} \max_{1 \leq j \leq J} \delta_S(\beta_j(\cdot)). \]
The desired result (50) follows from (51)-(55).
Inequality (2) follows from (41), (42), (46), and (50).

REFERENCES


32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.


43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?


56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.


64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.


71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.


77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.


82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.
95. CANCELLED
114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.
118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.
120. DONOHO, D. and LIU, R. (October 1987, revised March 1988). Geometrizing rates of convergence, II.
127. Same as No. 133.
128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.
131. Same as No. 140
163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform.
166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.
172. Adler, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.